Summary comments on analyses of the island closure experiment

D. S. Butterworth

Marine Resource Assessment and Management Group
Department of Mathematics and Applied Mathematics
University of Cape Town, Rondebosch, 7701

Summary

Based on a summary of issues raised in recently submitted documents, the conclusion is drawn that existing analyses based on annually-aggregated data are fully acceptable for consideration in developing management recommendations regarding possible future island closures., but that those based on individual data are not. The reason for this last statement is that results based on a methodology which an unchallenged proof has shown to be flawed are necessarily considered to be unreliable. It is also suggested that sufficient analyses of existing data using the estimation model of the 2016 Panel Algorithm are available to allow for proceeding towards formulation of recommendations. However, discussion is first needed on the response variables to be considered, reconciliation of apparent conflicts amongst some of these, and the criteria/considerations to be taken into account in developing those recommendations. In future, an important prerequisite will be a re-emphasis of the need to follow agreed protocols when engaged in a comparative analysis exercise such as this.

Analyses based on the Algorithm recommended by the Panel for the 2016 International Stock Assessment Workshop

Results of the most recent application of the estimation model of this algorithm (see Figure 1), developed in collaboration with and endorsed by IWS Panels, are reported in FISHERIES/2020/JAN/SWG-PEL/09rev for Dassen and Robben islands. The most recent estimates of the island closure effect parameter δ, based on the same default model, for St Croix and Bird islands are reported in MARAM/IWS/2019/PENG/P2 and FISHERIES/2019/NOV/SWG-PEL/33. These utilise the most recent data made available in terms of pre-agreed procedures to DEFF for such analyses, which are applied in terms of this algorithm to annually aggregated data.

The results from these analyses are consequently fully acceptable for consideration in developing management recommendations regarding possible future island closures.

Analyses by Sherley (and colleagues) based on the use of individual data

Sherley and colleagues have motivated this approach as providing greater precision and reliability (through achieving a greater number of data to estimable parameters ratio) for estimates of δ. However, such approaches may provide estimates of precision (e.g. standard errors – se's) for δ that are negatively biased because of the effects of pseudo-replication. This is of concern, because it could lead to an estimate of δ being considered to be reliably established as meaningful when this is not the case. Sherley and colleagues have attempted to address this concern by the use of estimation approaches incorporating random effects terms.

Two concerns have been raised concerning their approach. The first relates to the selection of the random effects structure used. The 2019 IWS Panel recommended a procedure for choosing the best such structure. This has been reasonably implemented for more recent results, simultaneously confirming the previous associated concern that estimation of δ is not robust to alternative selections. Earlier results reported using this approach, which failed to apply this selection procedure, are therefore confirmed to have been invalid.

Nevertheless, even when such a selection approach is incorporated, such random effects approaches cannot be guaranteed to fully account for pseudo replication effects (so may still yield negatively biased estimates of se's); but in this specific case there is a second and much more important concern (which has also been raised regularly, but has never received an adequate response, in the past). This concern is related to the structure of the data available from this experiment as this impacts the optimal precision which is achievable for estimates of δ. What is critical here is that the estimates of δ are informed by inter-annual changes in the data. For all response variables considered in the experiment, there is no linkage between elements of the individual data from one year to the next (e.g. there is no information collected that provides the ability to link a penguin or nest sampled one year to a sample taken the next year), so that the individual data are statistically independent from one year to the next. A little thought makes clear then that these individual data cannot add any further information content to the estimation of δ than is already contained in their annually aggregated value, and therefore cannot improve the estimation precision for δ. This contention has now been confirmed by what amounts to a mathematical-statistical proof (the Annex of FISHERIES/2020/AUG/SWG-PEL/82). Continued acceptance of results from this individual-based approach would therefore necessarily require that proof to be shown to be invalid.

Two recent contributions by Sherley serve to strengthen concerns about results from the individual-based approach. Comparisons in Table 1 of FISHERIES/2020/SEP/SWG-PEL/86 indicate in some cases much smaller standard errors for δ for individual- compared to aggregated approaches. Given the above, this makes clear that in those cases even though the random error structure selection procedure has been applied, it has been unable to account completely for pseudo-replication, hence providing false impressions of the precision of the result. In FISHERIES/2020/SEP/SWG-PEL/85 Sherley quotes Maunder (2001)[1] as a fisheries-related example of the equivalent of the two-step aggregated estimation approach leading to worse precision than a single step process (as in the individual-based approach). But Maunder (2001) failed to make any adjustment for pseudo-replication (the non-independence of his “individual-equivalent” data). In principle, that case could see potential utility for the Sherley individual-based approach, as those individual data involve vessels which are identifiable from one year to the next, and hence provide more inter-annual information content than annually aggregated values, unlike in this island closure case. However, this one-step process is generally not attempted in fisheries assessments for reasons which include that although random effects models may be used in the “standardization” process concerned, they are generally unable to account for all contributors to pseudo-replication effects. This necessitates a two-step process to estimate the size of “process error” (additional variance), as in the case of the Panel Algorithm of the previous section. Essentially, for this closure experiment as is typical in fisheries assessments, process error dominates observation error (MARAM/IWS/DEC15/PengD/P2).

No counter to the proof in FISHERIES/2020/AUG/SWG-PEL/82 has been offered by Sherley or his earlier co-authors. Estimates based on a methodology which an unchallenged proof has shown to be flawed are necessarily unreliable. The results from Sherley (and colleagues) based on their individual data-based analyses are consequently quite **un**acceptable for consideration in developing management recommendations regarding possible future island closures.

Possible further steps needed in moving towards management recommendations

Given the above, do the results provided in FISHERIES/2020/JAN/SWG-PEL/09rev for Dassen and Robben islands, and for St Croix and Bird islands as reported in MARAM/IWS/2019/PENG/P2 and FISHERIES/2019/NOV/SWG-PEL/33, already provide sufficient information on which to base recommendations, or are further analyses and/or is more discussion on elements thereof needed?

1) Data

Existing agreements are that results considered from analyses of the experiment are to be based on data submitted to DEFF and available to all analysts, and that data for all the years available should be used for default results (note that some of Sherley’s submissions have not respected all aspects of these agreements). In this respect then, there are sufficient analyses of existing data using the estimation model of the 2016 Panel Algorithm to proceed.

[1] CJFAS 58 (2001) 795-803

2) Co-variates to be considered

The default model for estimation of δ agreed at the 2015 IWS Panel meeting did not include co-variates. Sherley's models include species biomass estimates as co-variates. The 2015 Panel excluded those from the default, though suggesting such approaches as second priority sensitivities. Suggestions from penguin biologists on which co-variates might be used (e.g. for standardisation) have varied considerably over time (e.g. brood mass was suggested at the end of 2019, but not mentioned recently). Discussions are first needed to obtain clarification on such choices, as well as to check whether the associated data are available.

3) Response variables to be considered

In principle, chick survival is a valuable addition to the set of response variables for which data are available, as associated changes link directly to penguin population dynamics, and hence to estimates of related changes in population growth rates. However, the marked (and apparently relatively precisely estimated) change in the estimated survival rate at Robben (but not Dassen) island from the KM estimates of these rates after some 50 days exposure is a concern. A further concern is that although chick survival is a component of fledging success, for Dassen island these two response variables offer estimates of δ which have different signs and are (in simple terms) near statistically significantly different at the 10% level. Questions also arise about the reliability of the foraging-related variables given that the signs of point estimates of δ to which they give rise differ for the Western and the Eastern Cape islands concerned (see Figure 1). These are matters that require further discussion, as the reasons for the features above need to be better understood before the associated results can be used with confidence.

4) Linkage of response variables to penguin population dynamics

The 2015 IWS Panel stated clearly that a response variable should not be considered further if there was no (objective) way to link it to penguin population dynamics to allow changes in the variable to be linked to a threshold for the extent of improvement in penguin population growth rate. Only two of the current variables meet that criterion; chick survival and fledging success, but there are problems associated with both as outlined in the preceding paragraph. Penguin biologists were tasked to provide input on this for the other response variables, but as yet there has been no progress other than some suggestion for chick condition based on a relationship with penguin demographics established for a penguin population elsewhere in the world. At this time then, a more pragmatic approach is required which takes some form of account of δ estimates for the other response variables as well, possibly through appeal to some non-parametric statistics concepts. Again, this requires further discussion.

5) Limitation of range of hypotheses to be considered in a comparative exercise

Sherley (FISHERIES/2020/SEP/SWG-PEL/85) introduces the possibility of regime shifts in presenting arguments to explain the discrepancy between the chick survival and fledging success estimates of δ for Dassen island. But this approach is problematic if it is to be used *a posteriori* in this manner; especially because if such additions are to be allowed in a comparative exercise, the associated rules have to be agreed *a priori* – for example, such rules would need to cover acceptable criteria to justify the possibility of allowing for such assumptions, because in other years and for other variables such assumptions could change the conclusions to which their analyses would otherwise have led.

6) The process needed to compare results from different models in a comparative exercise

The 2015 Panel provided the default model to be used for providing estimates of δ. Certainly use of alternative approaches for estimation is desirable to check estimation robustness, but standardly in fisheries assessments this is required to be done through a "building-a-bridge" approach whereby factors that differ from the default are changed one at a time to enable an understanding, if there is a difference in results, of what aspect it is that is driving that difference. Compared to the agreed default approach, Sherley's results relate to applications that change many if not all of the following aspects: the data used, the period considered, inclusion of covariates, and working in normal rather than log space which leads to difficulties in relating his models' estimates of the closure effect parameter to those based on the default approach. There may be a case for some of these changes, but a comparative exercise is not assisted when the associated requirement to build-a-bridge is not followed. This becomes particularly relevant when Sherley claims that two independent sets of analyses have iterated to the stage where they are in effective agreement about impacts of fishing on penguins (FISHERIES/2020/JUL/SWG-PEL/53REV). The comparisons shown in Figure 1 show that this is hardly the case, with some important differences in the values and especially the variances for δ estimates readily evident. A further issue arises when Bayesian approaches, as in the case of Sherley's models, are being used. Such approaches require the specification of priors. For certain data, Sherley has been arguing that the closure effect parameter can be estimated as the same for both islands (e.g. FISHERIES/2020/JUL/SWG-PEL/53REV), under what is effectively the assumption of an uninformative prior. But that is a very questionable approach when there is clear evidence from other information (see Figure 1) that there **is** a real difference. Again, *a priori* discussions are needed on such matters.

7) Criteria for a basis to make recommendations for future closures to management

The PWG has been moving forward based on agreement with the Panel's recommendation for an objective decision rule to underpin such recommendations. In terms of estimation of the closure effect parameter δ (in log space), this effectively requires the point estimate to be less than -0.1. Sherley's recent documents (e.g. FISHERIES/2020/SEP/SWG-PEL/86) suggest moving away from this to consideration, for example, of Type I and Type II errors. To some extent though, such considerations have already been subsumed in the agreed value of the -0.1 threshold. Nevertheless, in the current situation which requires a somewhat pragmatic approach for the moment, further factors may need to be factored into a basis for recommendations, but this will first require further discussion.

In summary

The preceding comments in this section suggest the following actions at this time:

1) Sufficient analyses of existing data using the estimation model of the 2016 Panel Algorithm are now available to allow for proceeding towards the formulation of recommendations for management.

2) However, further discussion remains needed before such recommendations might be formulated on the response variables to be considered, reconciliation of apparent conflicts amongst some of these, and the criteria/considerations to be taken into account in developing those recommendations.

3) Other aspects can await "the next round", but an important prerequisite for that will be a re-emphasis on the need to follow agreed protocols when engaged in a comparative analysis exercise.

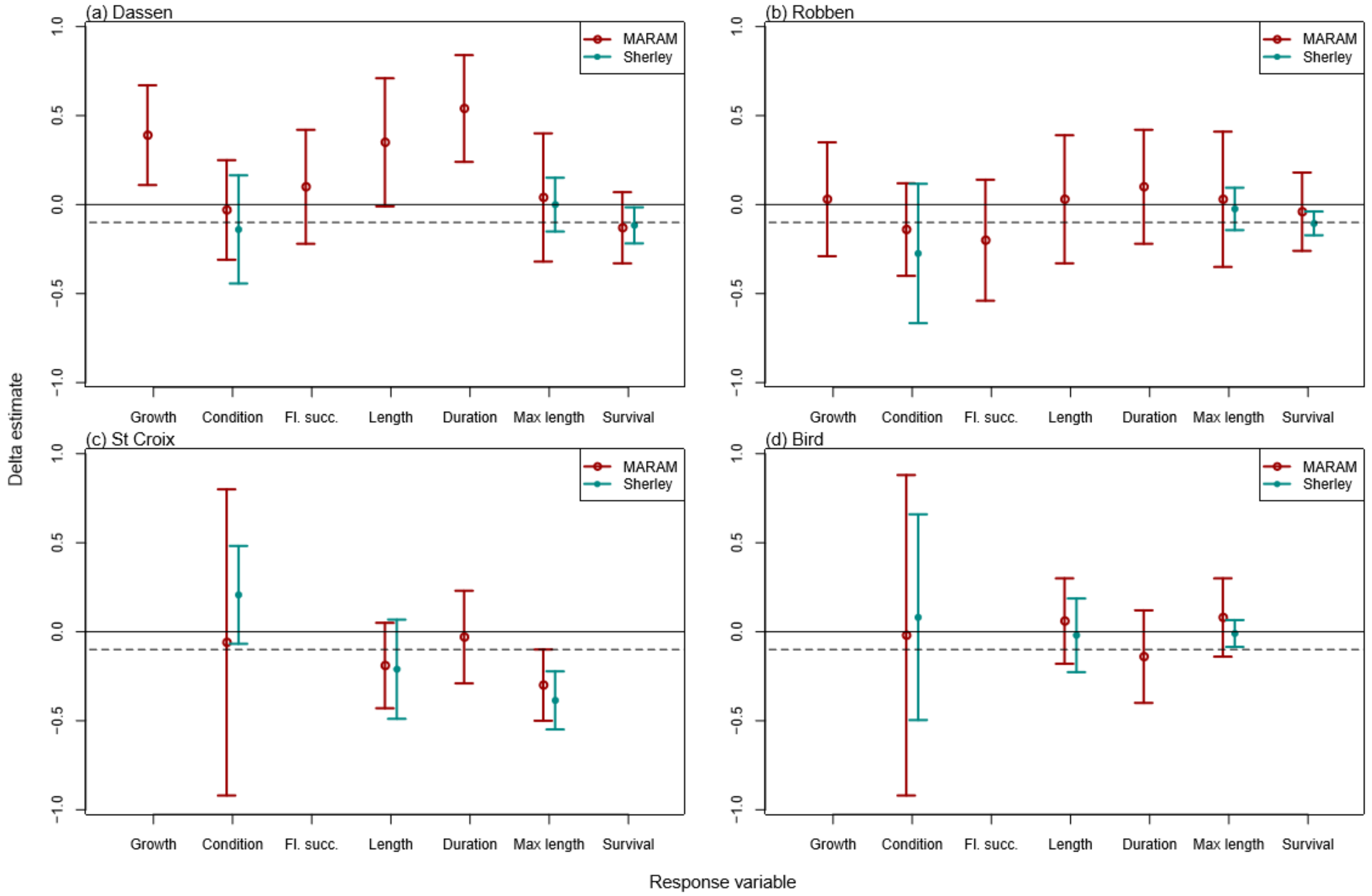

Figure 1: Zeh plots of the δ estimates and rough 95% confidence intervals are shown for the MARAM (aggregated data-based) and Sherley (individual data-based) models. The results for the MARAM models are taken from FISHERIES/2020/JAN/SWG-PEL/09rev for Robben and Dassen islands, from MARAM/IWS/2019/PENG/P2 for the foraging data for St Croix and Bird islands, and from FISHERIES/2019/NOV/SWG-PEL/33 for the chick condition data for St Croix and Bird islands. The values for the Sherley models have been derived from the last table of FISHERIES/2020/SEP/SWG-PEL/95 by use the following formula: $\delta = \ln(1 - p/100)$ where the p values are those reported in that last table as a simple approach to transform from normal to log-space to achieve improved comparability. The confidence intervals have been converted in a similar manner, and a rough standard error may be calculated as (max(CI)-min(CI))/4.. The Figure has been kindly provided by A. Ross-Gillespie.